(51) International Patent Classification 6 : C12N 15/10, C12Q 1/68	A1	(11) International Publication Number: WO 98/41623 (43) International Publication Date: 24 September 1998 (24.09.98)

(54) Title: SHUFFLING OF HETEROLOGOUS DNA SEQUENCES

(57) Abstract

The present invention relates to a new method of shuffling especially heterologous polynucleotide sequences, screening and/or selection of new recombinant proteins resulting therefrom having a desired biological activity, and especially to production and identification of novel proteases exhibiting desired properties. The method comprises the following steps: i) identification of at least one conserved region between the heterologous sequences of interest; ii) generating fragments of each of the heterologous sequences of interest, wherein said fragments comprise the conserved region(s), in a preferred embodiment due to the use of parts of the regions(s) as primers; and iii) shuffling/recombining said fragments using the conserved region(s) as (a) homologous linking point(s).

TITLE: Shuffling of heterologous DNA sequences

FIELD OF THE INVENTION

The present invention relates to a new method of shuffling
5 especially heterologous polynucleotide sequences, screening and/or
selection of new recombinant proteins resulting therefrom having a
desired biological activity, and especially to the production and
identification of novel proteases exhibiting desired properties.

10 BACKGROUND OF THE INVENTION

It is generally found that a protein performing a certain
bioactivity exhibits a certain variation between genera, and even
between members of the same species differences may exist. This
variation is even more outspoken at the genomic level.

15   This natural genetic diversity among genes coding for pro-
teins having basically the same bioactivity has been generated in
nature over billions of years and reflects a natural optimisation
of the proteins coded for in respect of the environment of the or-
ganism in question.

20   However, in general it has been found that the naturally oc-
curring bioactive molecules are not optimized for the various uses
to which they are put by mankind, especially when they are used
for industrial purposes.

It has therefore been of interest to industry to identify
25 such bioactive proteins that exhibit optimal properties in respect
of the use for which it is intended.

This has been done for many years by screening of natural
sources, or by use of mutagenesis. For instance, within the tech-
nical field of enzymes for use in e.g. detergents, the washing
30 and/or dishwashing performance of e.g. naturally occurring prote-
ases, lipases, amylases and cellulases has been improved signifi-
cantly by in vitro modifications of the enzymes.

In most cases these improvements have been obtained by site-
directed mutagenesis resulting in substitution, deletion or inser-
35 tion of specific amino acid residues which have been chosen either
on the basis of their type or on the basis of their location in
the secondary or tertiary structure of the mature enzyme (see for
instance US patent no. 4,518,584).

<u>Prior Art:</u>

Numerous methods to create genetic diversity, such as by
site directed or random mutagenesis, have been proposed and de-
5 scribed in scientific literature as well as patent applications.
For further details in this respect reference is made to the re-
lated art section of WO 95/22625, wherein a review is provided.

One method of the shuffling of homologous DNA sequences has
been described by Stemmer (Stemmer, (1994), **Proc. Natl. Acad. Sci.
10 USA**, Vol. 91, 10747-10751; Stemmer, (1994), **Nature**, vol. 370, 389-
391). The method concerns shuffling homologous DNA sequences by
using *in vitro* PCR techniques. Positive recombinant genes contain-
ing shuffled DNA sequences are selected from a DNA library based
on the improved function of the expressed proteins.

15   WO 95/22625 is believed to be the most pertinent reference
in relation to the present invention in its "gene shuffling" as-
pect. In WO 95/22625 a method for shuffling of homologous DNA se-
quences is described. An important step in the method described in
WO 95/22625 is to cleave the homologous template double-stranded
20 polynucleotide into random fragments of a desired size followed by
homologously reassembling of the fragments into full-length genes.

A disadvantage inherent to the method of WO 95/22625 is,
however, that the diversity generated through that method is lim-
ited due to the use of homologous gene sequences (as defined in WO
25 95/22625).

Another disadvantage in the method of WO 95/22625 lies in
the production of the random fragments by the cleavage of the tem-
plate double-stranded polynucleotide.

A further reference of interest is WO 95/17413 describing a
30 method of gene or DNA shuffling by recombination of DNA sequences
either by recombination of synthesized double-stranded fragments
or recombination of PCR generated sequences. According to the
method described in WO 95/17413 the recombination has to be per-
formed among DNA sequences with sufficient sequence homology to
35 enable hybridization of the different sequences to be recombined.

WO 95/17413 therefore also entails the disadvantage that the
diversity generated is relatively limited.

3

The present invention does not contain any steps involving production of random fragments by the cleavage of the template double-stranded polynucleotide, as described in WO 95/22625.

Further, WO 95/22625 relates to shuffling of homologous
5 genes, while the present invention relates to shuffling of heterologous genes.

SUMMARY OF THE INVENTION:

The problem to be solved by the present invention is to
10 avoid the limitation of shuffling only homologous DNA sequences by providing a method to shuffle/recombine heterologous sequences of interest.

The solution is to use at least one "conserved sequence region", wherein there is a sufficient degree of homology between
15 the heterologous sequences to be shuffled, as a "linking point" between said heterologous sequences.

Accordingly, a first aspect of the invention relates to a method of shuffling of heterologous sequences of interest comprising the following steps,

20   i)   identification of at least one conserved region between the heterologous sequences of interest;

        ii)   generating fragments of each of the heterologous sequences of interest, wherein said fragments comprise the conserved region(s); and

25   iii) shuffling/recombining said fragments using the conserved region(s) as (a) homologous linking point(s).

In an second aspect the invention relates to a method for producing a shuffled protein having a desired biological activity
30 comprising in addition to the steps of the first aspect the further steps:

        iv)   expressing the numerous different recombinant proteins encoded by the numerous different shuffled sequences from step iii); and

35   v)   screen or select the numerous different recombinant proteins from step ii) in a suitable screening or selection system for one or more recombinant protein(s) having a desired activity.

The term "conserved region" denotes a sequence region (preferably of at least 10 bp), wherein there is a relatively high sequence identity between said heterologous sequences.

In order for the conserved region to be used as "linking point" between said heterologous sequences, the sequence identity between the heterologous sequences, within said conserved regions, is sufficiently high to enable hybridization of the heterologous sequences using said conserved region as hybridization point ("linking point").

BRIEF DECRIPTION OF DRAWINGS

Fig. 1: Fig 1 illustrates the general concept of the invention, where

a) the black boxes define mutual, common, conserved regions of the sequences of interest, and

b) the PCR primers named "a,a´,b,b´,etc.." are primers directed to the conserved regions. Primers ("a´" and "b"), ("b´" and "c") etc.. have a sequence overlap of preferably at least 10 bp, and

c) primers "z" and "z´" are primers directed to the flanking parts of the sequence area of the sequences of interest which are shuffled according to the method of the invention.

Fig 2: Shows an alignment of 5 protease (subtilase) DNA sequences. Herein are a number of conserved regions such as the common partial sequences numbered 1-5.

Fig 3: Shows an alignment of different lipases.

DEFINITIONS

Prior to discussing this invention in further detail, the following terms will be defined.

"Shuffling": The term "shuffling" means recombination of nucleotide sequence(s) between two or more DNA sequences of interest resulting in output DNA sequences (i.e. DNA sequences having been subjected to a shuffling cycle) having a number of nucleo-

tides exchanged, in comparison to the input DNA sequences (i.e. starting point DNA sequences of interest).

Alternatively, the term "shuffling" may be termed "recombining".

5 "Homology of DNA sequences": In the present context the degree of DNA sequence homology is determined as the degree of identity between two sequences indicating a derivation of the first sequence from the second. The homology may suitably be determined by means of computer programs known in the art, such as GAP provided in the GCG program package (Program Manual for the Wisconsin

10 Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711)(Needleman, S.B. and Wunsch, C.D., (1970), Journal of Molecular Biology, 48, 443-453).

"Homologous": The term "homologous" means that one single-

15 stranded nucleic acid sequence may hybridize to a complementary single-stranded nucleic acid sequence. The degree of hybridization may depend on a number of factors including the amount of identity between the sequences and the hybridization conditions such as temperature and salt concentration as discussed later

20 (vide infra).

Using the computer program GAP (vide supra) with the following settings for DNA sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, it is in the present context believed that two DNA sequences will be able to hybridize (using

25 medium stringency hybridization conditions as defined below) if they mutually exhibit a degree of identity of at least 50%, more preferably at least 60%, more preferably at least 70%, more preferably at least 80%, more preferably at least 85%, and even more preferably at least 90%.

30 "Heterologous": Two DNA sequences are said to be heterologous if one of them comprises a partial sequence of at least 40 bp which does not exhibit a degree of identity of more than 50%, more preferably of more than 70%, more preferably of more than 80%, more preferably of more than 85%, more preferably of more than

35 90%, and even more preferably of more than 95%, of any partial sequence in the other. More preferably the first partial sequence is at least 60 bp, more preferably the first partial sequence is at least 80 bp, even more preferably the first partial sequence is at

least 120 bp, and most preferably the first partial sequence is at least 500 bp.

"Hybridization:" Suitable experimental conditions for determining if two or more DNA sequences of interest do hybridize or not are herein defined as hybridization at medium stringency as described in detail below.

A suitable experimental low stringency hybridization protocol between two DNA sequences of interest involves presoaking of a filter containing the DNA fragments to hybridize in 5 x SSC (Sodium chloride/Sodium citrate, Sambrook et al. 1989) for 10 min, and prehybridization of the filter in a solution of 5 x SSC, 5 x Denhardt's solution (Sambrook et al. 1989), 0.5 % SDS and 100 µg/ml of denatured sonicated salmon sperm DNA (Sambrook et al. 1989), followed by hybridization in the same solution containing a concentration of 10ng/ml of a random-primed (Feinberg, A. P. and Vogelstein, B. (1983) Anal. Biochem. 132:6-13), 32P-dCTP-labeled (specific activity > 1 x 109 cpm/µg) probe (DNA sequence) for 12 hours at approx. 45°C. The filter is then washed twice for 30 minutes in 2 x SSC, 0.5 % SDS at least 55°C, more preferably at least 60°C, and even more preferably at least 65°C (high stringency).

Molecules to which the oligonucleotide probe hybridizes under these conditions are detected using an X-ray film.

"Alignment": The term "alignment" used herein in connection with an alignment of a number of DNA and/or amino acid sequences means that the sequences of interest are aligned in order to identify mutual/common sequences of homology/identity between the sequences of interest. This procedure is used to identify common "conserved regions" (vide infra), between sequences of interest. An alignment may suitably be determined by means of computer programs known in the art, such as PILEUP provided in the GCG program package (Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711)(Needleman, S.B. and Wunsch, C.D., (1970), Journal of Molecular Biology, 48, 443-453).

"Conserved regions:" The term "conserved region" used herein in connection with a "conserved region" between DNA and/or amino acid sequences of interest means a mutual, common sequence region

of two or more sequences of interest, wherein there is a relatively high degree of sequence identity between two or more of the heterologous sequences of interest. In the present context a conserved region is preferably at least 10 base pairs (bp), more
5 preferably at least 20 bp, and even more preferably at least 30 bp.

Using the computer program GAP (*vide supra*) with the following settings for DNA sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the degree of DNA sequence
10 identity within the conserved region, between two or more of the heterologous sequences of interest, is preferably at least 80%, more preferably at least 85%, more preferably at least 90%, and even more preferably at least 95%.

"Primer": The term "primer" used herein, especially in con-
15 nection with a PCR reaction, is a primer (especially a "PCR-primer") defined and constructed according to general standard specification known in the art ("PCR A practical approach" IRL Press, (1991)).

"A primer directed to a sequence:" The term "a primer di-
20 rected to a sequence" means that the primer (preferably to be used in a PCR reaction) is constructed so as to exhibit at least 80% degree of sequence identity to the sequence part of interest, more preferably at least 90% degree of sequence identity to the sequence part of interest, which said primer consequently is
25 "directed to".

"Sequence overlap extension PCR reaction (SOE-PCR)": The term "SOE-PCR" is a standard PCR reaction protocol known in the art, and in the present context it is defined and performed according to standard protocols defined in the art ("PCR A practical
30 approach" IRL Press, (1991)).

"Flanking": The term "flanking" used herein in connection with DNA sequences comprised in a PCR-fragment means the outmost end partial sequences of the PCR-fragment, both in the 5` and 3` ends of the PCR fragment.

35   "Subtilases": A serine protease is an enzyme which catalyzes the hydrolysis of peptide bonds, and in which there is an essential serine residue at the active site (White, Handler and Smith,

8

1973 "Principles of Biochemistry," Fifth Edition, McGraw-Hill Book
Company, NY, pp. 271-272).

The bacterial serine proteases have molecular weights in the
range of 20,000 to 45,000 Daltons. They are inhibited by diisopro-
5 pylfluorophosphate. They hydrolyze simple terminal esters and are
similar in activity to eukaryotic chymotrypsin, also a serine pro-
tease. A more narrow term, alkaline protease, covering a
sub-group, reflects the high pH optimum of some of the serine pro-
teases, from pH 9.0 to 11.0 (for review, see Priest (1977)
10 Bacteriological Rev. 41 711-753).

A sub-group of the serine proteases tentatively designated
subtilases has been proposed by Siezen et al., Protein Engng. 4
(1991) 719-737. They are defined by homology analysis of more than
40 amino acid sequences of serine proteases previously referred to
15 as subtilisin-like proteases.

DETAILED DESCRIPTION OF THE INVENTION
A method for shuffling heterologous sequences of interest

In a preferred embodiment the fragments generated in step
20 ii) of the first aspect of the invention is generated by use of
PCR technology.

Accordingly, an aspect of the invention relates to a method
of shuffling of heterologous DNA sequences of interest, according
to the first aspect of the invention, comprising the following
25 steps

i)   identification of one or more conserved region(s) (hereafter
        named "A,B,C" etc..) in two or more of the heterologous
        sequences;

ii)   construction of at least two sets of PCR primers (each set
30   comprising a sense and an anti-sense primer) for one or more
        conserved region(s) identified in i) wherein

            in one set the sense primer (named: "a"=sense primer)
            is directed to a sequence region 5´ (sense strand) of
            said conserved region (e.g. conserved region "A"), and
35   the anti-sense primer (named "a´"=anti-sense primer)
            is directed either to a sequence region 3´ (sense
            strand) of said conserved region or directed to a

9

sequence region at least partially within said
conserved region,

and in another set the sense primer (named: "b"=sense
primer) is directed either to a sequence region 5´
5   (sense strand) of said conserved region or directed to
a sequence region at least partially within said
conserved region and the anti-sense primer (named:
"b´"=anti-sense primer) is directed to a sequence
region 3´ (sense strand) of said conserved region
10   (e.g. conserved region "A"), and

the two sequence regions defined by the regions
between primer set "a" and "a´" and "b" and "b´" (both
said regions is including the actual primer sequences)
have a homologous sequence overlap of at least 10 base
15   pairs (bp) within the conserved region;

iii)   for one or more identified conserved regions of interest in
step i) two PCR amplification reactions are performed with
the heterologous DNA sequences in step i) as template, and
where

20   one of the PCR reactions uses the 5´ primer set
identified in step ii) (e.g. named "a","a´") and the
second PCR reaction uses the 3´ primer set identified
in step ii) (e.g. named "b","b´");

iv)   isolation of the PCR fragments generated as described in
25   step iii) for one or more of the identified conserved region
in step i);

v)   pooling of two or more isolated PCR fragments from step iv)
and performing a Sequence overlap extension PCR reaction
(SOE-PCR) using said isolated PCR fragments as templates;
30   and

vi)   isolation of the PCR fragment obtained in step v), wherein
said isolated PCR fragment comprises numerous different
shuffled sequences containing a shuffled mixture of the PCR
fragments isolated in step iv), wherein said shuffled
35   sequences are

characterized in that the partial DNA sequences, originating from
the homologous sequence overlaps in step ii), have at least 80%

identity to one or more partial sequences in one or more of the
original heterologous DNA sequences in step i).

A method of producing one or more recombinant protein(s) having a
5 desired biological activity

        In an second aspect the invention relates to a method of
producing a shuffled protein having a desired biological activity
comprising in addition to the steps i) to vi) immediately above
the further steps:
10 vii)   expressing the numerous different recombinant proteins
             encoded by the numerous different shuffled sequences
             in step vi); and

    viii)   screen or select the numerous different recombinant
             proteins from step vii) in a suitable screening or se-
15   lection system for one or more recombinant protein(s)
             having a desired activity.

Heterologous DNA sequences

        The method of the present invention may be used to shuffle
20 basically all heterologous DNA sequences of interest.

        Preferably, it is used to shuffle heterologous DNA sequences
encoding an enzymatic activity, such as amylase, lipase, cutinase,
cellulase, oxidase, phytase, and protease activity.

        An further advantage of the present method is that it makes
25 it possible to shuffle heterologous sequences encoding different
activities, e.g. different enzymatic activities.

        The method of the invention is in particular suitable to
shuffle heterologous DNA sequences encoding a protease activity,
in particular a subtilase activity.

30   A number of subtilase DNA sequences are published in the
art. A number of those subtilase DNA sequences are in the present
context heterologous DNA sequences, and it is generally believed
that they are mutually too heterologous to be shuffled by the
shuffling methods presently known in the art (WO 95/17413, WO
35 95/22625). However the method according to the invention enables
shuffling of such sequences. For further details reference is made
to a working example herein (vide infra).

11

Further, the present invention is suitable to shuffle dif-
ferent lipase sequences. For further details reference is made to
a working example herein (*vide infra*).

The heterologous DNA sequences used as templates may before-
5 hand have been cloned into suitable vectors, such as a plasmid.
Alternatively, a PCR-reaction may be performed directly on micro-
organisms known to comprise the DNA sequence of interest according
to standard PCR protocols known in the art.

10 Identification of one or more conserved regions in heterologous
sequences:

Identification of conserved regions may be done by an align-
ment of the heterologous sequences by standard computer programs
(*vide supra*).

15   Alternatively, the method may be performed on completely new
sequences, where the relevant "conserved regions" are chosen as
conserved regions which are known in the art to be conserved re-
gions for this particular class of proteins.

E.g., the method may be used to shuffle completely unknown
20 subtilase sequences, which are known to be very conserved in e.g.
regions around the active site amino acids. PCR reaction may then
be performed directly on new unknown strains with primers directed
to those conserved regions.

25 PCR-primers

The PCR primers are constructed according to the standard
descriptions in the art. Preferably, they are 10-75 base pairs
(bp) long.

30 Homologous sequence overlap

In step ii) of claim 3 of the invention the two sequence
regions defined by the regions between primer set "a" and "a'" and
"b" and "b'" (both said regions is including the actual primer
sequences) have a homologous sequence overlap of at least 10 base
35 pairs (bp) within the conserved region.

Said homologous sequence overlap is more preferably of at
least 15 bp, more preferably of at least 20 bp, and even more
preferably of at least 35 bp.

   The homologous sequence overlaps in step ii) of claim 3 have
at least 80% identity to one or more partial sequences in one or
more of the original heterologous DNA sequences in step i) of said
claim, more preferably the homologous sequence overlaps in step
5 ii) have at least 90% identity to one or more partial sequences in
one or more of the original heterologous DNA sequences in step i)
of said claim, and even more preferably the homologous sequence
overlaps in step ii) have at least 95% identity to one or more
partial sequences in one or more of the original heterologous DNA
10 sequences in step i) of said claim.

PCR-reactions

   If not otherwise mentioned the PCR-reaction performed ac-
cording to the invention is performed according to standard proto-
15 cols known in the art.

   The term "Isolation of PCR fragment" is intended to cover an
aliquot containing the PCR fragment. However, the PCR fragment is
preferably isolated to an extent which removes surplus of primers,
nucleotides, etc.

20   Further, the fragment used for SOE-PCR in step v) of claim
3, may alternatively be generated by other processes than the PCR
amplification process described in step iii) of said claim. Suit-
able fragments used for the SOE-PCR in step v), may e.g. be gener-
ated by cutting out suitable fragments by restriction enzyme di-
25 gestion at appropriate sites (e.g. restriction sites situated on
each site of a conserved region identified in step i). Such alter-
native processes for generating such suitable fragments for use in
the SOE-PCR in step v) are considered within the scope of the in-
vention.

30   In an embodiment of the invention the PCR DNA fragment(s)
is(are) prepared under conditions resulting in a low, medium or
high random mutagenesis frequency.

   To obtain low mutagenesis frequency the DNA sequence(s)
(comprising the DNA fragment(s)) may be prepared by a standard PCR
35 amplification method (US 4,683,202 or Saiki et al., (1988), Sci-
ence 239, 487 - 491).

   A medium or high mutagenesis frequency may be obtained by
performing the PCR amplification under conditions which increase

the misincorporation of nucleotides, for instance as described by
Deshler, (1992), GATA 9(4), 103-106; Leung et al., (1989), Tech-
nique, Vol. 1, No. 1, 11-15.

5   Final shuffles sequences

        One of the advantages of the present invention is that the
final "shuffled sequences" in step vi) of claim 3 of the present
invention only comprise sequence information which is originally
derived from the original heterologous sequences of interest in
10   step i) of said claim. The present invention does not use artifi-
cially made "linker sequences" to recombine one or more of the
heterologous sequences, which is a strategy known in the art to
e.g. be able to shuffle different domains in proteins, wherein
each domain is encoded by different heterologous sequences (WO
15   95/17413).

        Accordingly, the invention relates to a method characterized
in that each of the shuffled sequences, the partial DNA sequences,
originating from the homologous sequence overlaps in step ii),
only contains sequence information which is originally derived
20   from the original heterologous sequences in step i) (in the first
to third aspect of the invention) (i.e. said "homologous sequence
overlaps" in step ii) has at least 80% identity to one or more
partial sequences in one or more of the original heterologous DNA
sequences in step i).

25   More preferably, the "homologous sequence overlaps" in step
ii) have at least 90% identity to one or more partial sequences in
one or more of the original heterologous DNA sequences in step i);
and even more preferably the "homologous sequence overlaps" in
step ii) have at least 95% identity to one or more partial se-
30   quences in one or more of the original heterologous DNA sequences
in step i), and most preferably the "homologous sequence overlaps"
in step ii) have 100% identity to one or more partial sequences in
one or more of the original heterologous DNA sequences in step i).

35   Expressing the recombinant protein from the shuffled sequences

        Expression of the recombinant protein encoded by the shuf-
fled sequence of the present invention may be performed by use of

14

standard expression vectors and corresponding expression systems
known in the art.

Suitable screening or selection system

5   In its second aspect the present invention relates to a
method for producing one or more recombinant protein(s) having a
desired biological activity.
        A suitable screening or selection system will depend on the
desired biological activity.
10   A number of suitable screening or selection systems to
screen or select for a desired biological activity are described
in the art. Examples are:
        Strauberg et al. (Biotechnology 13: 669-673 (1995), which
describes a screening system to screen for subtilisin variants
15 having a calcium-independent stability;
        Bryan et al. (Proteins 1:326-334 (1986)), which describes a
screening assay to screen for proteases having enhanced thermal
stability; and
        WO 97/04079 which describes a screening assay to screen for
20 lipases having an improved wash performance in washing detergents.
        A preferred embodiment of the invention comprises screening
or selection of recombinant protein(s), wherein the desired bio-
logical activity is performance in dish-washing or laundry deter-
gents. Examples of suitable dish-washing or laundry detergents are
25 disclosed in WO 97/04079 and WO 95/30011.
        The invention is described in further detail in the follow-
ing examples which are not in any way intended to limit the scope
of the invention.

30 MATERIALS AND METHODS

Strains
    E. coli strain: DH10B (Life Technologies)

35 Bacillus subtilis strain: DN1885 amyE. A derivative of B,s
    168RUB200 (J. Bacteriology 172:4315-4321 (1990))

15

Plasmids

pKH400: pKH400 was constructed from pJS3 (E. coli - B. subtilis
shuttle vector containing a synthetic gene encoding for subtilase
309 (described by Jacob Schiodt et al. in Protein and Peptide
5 letters ?:39-44 (1996)), by introduction of two BamHI sites at
positions 1841 and 3992.

Protease sequences used for shuffling

GenBank entries A13050_1, D26542, A22550, Swiss-Prot entry
10 SUBT_BACAM P00782, and PD498 (Patent Application No. WO 96/34963).

General molecular biology methods

Unless otherwise mentioned the DNA manipulations and
transformations were performed using standard methods of molecular
15 biology (Sambrook et al. (1989) Molecular cloning: A laboratory
manual, Cold Spring Harbor lab., Cold Spring Harbor, NY; Ausubel,
F. M. et al. (eds.) "Current protocols in Molecular Biology". John
Wiley and Sons, 1995; Harwood, C. R., and Cutting, S. M. (eds.)
"Molecular Biological Methods for Bacillus". John Wiley and Sons,
20 1990).

        Enzymes for DNA manipulations were used according to the
specifications of the suppliers.

Enzymes for DNA manipulations

25 Unless otherwise mentioned all enzymes for DNA manipulations, such
as e.g. restiction endonucleases, ligases etc., are obtained from
New England Biolabs, Inc.

EXAMPLES

30

EXAMPLE 1

A) Vector construction

35 1) Amplification of the pre-pro sequences

        Host cells harboring the plasmid DNA encoding the full length
enzymes A13050_1 (GenBank), SUBT_BACAM P00782 (Swiss-Prot), D26542
(GenBank), A22550 (GenBank), and PD498 (Patent Application No. WO

96/34963) were starting material. By standard mini-prep isolation
of plasmid DNA, purified DNA was obtained. With these template
DNAs, 5 standard PCRs were performed to amplify the respective
pre-pro sequences. The fragments were generated using the proof
reading Pwo DNA polymerase (Boehringer Mannheim) and the following
sets of primers directed against the very N- and C-termini of the
respective pre-pro sequences:

A13050_1
TiK111: 5' GAG GAG GGA AAC CGA ATG AGG AAA AAG AGT TTT TGG.
TiK117: 5' CGC GGT CGG GTA CCG TTT GCG CCA AGG CAT G.

SUBT_BACAM P00782
TiK112: 5' GAG GAG GGA AAC CGA ATG AGA GGC AAA AAA GTA TGG.
TiK118: 5' CGC GGT CGG GTA CCG ACT GCG CGT ACG CAT G.

D26542
TiK110: 5' GAG GAG GGA AAC CGA ATG AGA CAA AGT CTA AAA GTT ATG.
TiK116: 5' CGC GGT CGG GTA CCG TTT GAC TGA TGG TTA CTT C.

A22550
TiK109: 5' GAG GAG GGA AAC CGA ATG AAG AAA CCG TTG GGG.
TiK115: 5' CGC GGT CGG GTA CCG ATT GCG CCA TTG TCG TTA C.

PD498
TiK113: 5' GAG GAG GGA AAC CGA ATG AAG TTC AAA AAA ATA GCC.
TiK119: 5' CGC GGT CGG GTA CCG CAG AAT AGT AAG GGT CAT TC.

    The obtained DNA fragments of a length between 300-400 bp
were purified by agarose gel-electrophoresis with subsequent gel
extraction (QIAGEN) and subjected to assembly by splice-by-overlap
extension PCR (SOE-PCR).

    2) SOE-PCR
    The pre-pro fragments were then separately spliced by SOE-PCR
to the 3' part of the promoter of the vector pKH400. The 3' part
of the promoter was obtained by standard PCR with the Pwo DNA po-
lymerase using 1 ng of pKH400 as template and the primers:

17

TiK106: 5' CGA CGG CCA GCA TTG G.

TiK107: 5' CAT TCG GTT TCC CTC CTC.

The resulting 160 bp fragment was gel-purified. Subsequently, 5
SOE-PCRs were performed under standard conditions (Pwo DNA po-
5 lymerase) using as template each of the 5 pre-pro sequences mixed
with equal molar amounts of the 3' part of the promoter. The as-
sembling primers were:

TiK120: 5' CTT TGA TAC GTT TAA ACT ACC.

TiK121: 5' CGC GGT CGG GTA CCG.

10 The obtained fragments were also gel-purified.

3) Insertion of the pre-pro sequences into the pKH400 shuttle vec-
tor

The pKH400 vector was cut with Pme I and Acc65 I to remove
15 the existing linker sequence. The 5 purified SOE-PCR fragments
from 2) were also digested with the same enzymes and gel-purified.
Only with the SOE-PCR of the SUBT_BACAM P00782 pre-pro sequence
special caution was required because it contains an internal Pme
I-site so that a partial digest was performed. In separate stan-
20 dard ligation mixes the pre-pro fragments were then ligated to the
pKH400 vector. After transformation of DH10B E.coli cells, colo-
nies were selected on ampicillin containing media. Correctly
transformed cells were identified by control digest and sequenced.
The thus obtained vectors were named pTK4001-4005.

25

B) Preparation of the small fragments of the proteases A13050_1
(GenBank), SUBT_BACAM P00782 (Swiss-Prot), D26542 (GenBank),
A22550 (GenBank), and PD498 (Patent Application No. WO 96/34963).

30 1) Standard PCR reactions were assembled with 0.5 μl of mini-prep
DNA of each protease gene as templates. Since these five protease
genes shall be fragmented into six fragments (I-VI), 30 PCRs are
required (see fig 1). The Ampli-Taq polymerase (5U) was used in
combination with the following primer sets (the numbering corre-
35 sponds to the amino acid position in A22550). If there are primers
labeled #.1, #.2, etc., then equal molar amounts of them are mixed
prior to PCR and treated as one primer in the PCR:

Set I)

TiK122.1 (116-124)

5' CCG GCG CAG GCG GTA CCX TRS GGX ATW XCX CXX RTX MAA GC.

TiK122.2 (116-124)

5 5' CCG GCG CAG GCG GTA CCX TRS GGX ATW XCA WWC ATX WAT AC.

TiK123 (174-180)

5' GTT CCX GCX ACR TGX GTX CC.

Set II)

10 TiK124 (174-180)

5' GGX ACX CAY GTX GCX GGA AC.

TiK125.1 (217-223)

5' GCC CAC TSX AKX CCG YTX AC.

TiK125.2 (217-223)

15 5' GCC CAC TSX AKX CCT YGX GC.

TiK125.3 (217-223)

5' GCC CAX TSR AKX CCK XXX RCW AT.

Set III)

20 TiK126.1 (217-223)

5' GTX ARC GGX MTX SAG TGG GC.

TiK126.2 (217-223)

5' GCX CRA GGX MTX SAG TGG GC.

TiK126.3 (217-223)

25 5' TWG CYC AAG GWW TXS AXT GKR.

TiK126.5 (217-223)

5' TWG CTC AAG GHH THS ART GG.

TiK127.1 (255-261)

5' GCX GCX ACX ACX ASX ACX CC.

30 TiK127.2 (255-261)

5' GCY SCW AYW AMX AGW AYA YCA.

Set IV)

TiK128.1 (255-261)

35 5' GGX GTX STX GTX GTX GCX GC.

TiK128.2 (255-261)

5' TGR TRT WCT MKT WRT WGS RGC.

TiK129.1 (292-299)

5' GBX CCX ACR YTX GAR AAW GAX G.

TiK129.2 (292-299)

5' GBX CCR TAC TGX GAR AAR CTX G.

TiK129.3 (292-299)

5 5' GKX CCA TAC KKA GAR AAR YTT G.

TiK129.5 (292-299)

5' GKR CCA TAC KKA GAR AAG YTT G.

Set V)

10 TiK130.1 (292-299)

5' CXT CWT TYT CXA RYG TXG GXV C.

TiK130.2 (292-299)

5' CXA GYT TYT CXC AGT AYG GXV C.

TiK130.3 (292-299)

15 5' CAA GYT TCT CTM MGT ATG GSM C.

TiK130.5 (292-299)

5' CAA GTT TCT CTC AGT ATG GGA C.

TiK131.1 (324-330)

5' GGX GWX GCC ATX GAY GTX CC.

20 TiK131.2 (324-330)

5' GGA GTA GCC ATX GAX GTW CC.

Set VI)

TiK132.1 (324-330)

25 5' GGX ACR TCX ATG GCX WCX CC.

TiK132.2 (324-330)

5' GGW ACX TCX ATG GCA WCX CC.

TiK133.1 (375-380)

5' CGG CCC CGA CGC GTT TAC YGX RYX GCX SYY TSX RC.

30 TiK133.2 (375-380)

5' CGG CCC CGA CGC GTT TAT CXT RYX GCX XXY TYW G.

TiK133.3 (375-380)

5' CGG CCC CGA CGC GTT TAT CXT RCX GCX GCX TYT GMR TT.

TiK133.4 (375-380)

35 5' CGG CCC CGA CGC GTT TAT CTT ACG GCA GCC TCA GC.

(X = deoxy-inosine, Y = 50% C + 50% T, R = 50% A + 50% G, S = 50%

C + 50% G, W = 50% A + 50% T, K = 50% T + 50% G, M = 50% A + 50%

20

C, B = 33.3% C + 33.3% G + 33.3% T, V = 33.3% C + 33.3% G + 33.3% A, H = 33.3% C + 33.3% A + 33.3%).

After 30 cycles at annealing temperatures ranging from 40-
5 60°C the amplified fragments were gel-purified and recovered.

2) SOE-PCR to randomly assemble the small fragments
     Equimolar amounts of each of the purified fragments were taken and mixed in one tube as templates for assembly in an otherwise standard SOE-PCR with Ampli-Taq polymerase. The external
10 primers used are:
          TiK134.1:   CCG GCG CAG GCG GTA CC.
          TiK135.1:   CGG CCC CGA CGC GTT TA.

15 Also the primer pairs
          TiK134.2:   GGC GCA GGC GGT AC.
          TiK135.2:   GCC CCG ACG CGT TTA.
     and
          TiK134.3:   CGC AGG CGG TAC.
20   TiK135.3:   CCC GAC GCG TT.
     can be used. The annealing temperatures are ranging from 40°C to 70°C.
     The re-assembly is also achieved by sequentially re-assembling all conceivable combinations of fragments, e.g.: In
25 tube 1 all seven fragments obtained by PCR with the primers of set I (see above, B1-2) are mixed, in tube 2 fragments obtained by PCR with the primers of set II are mixed, in tube 3 fragments obtained by PCR with the primers of set III are mixed, in tube 4 fragments obtained by PCR with the primers of set IV are mixed, in tube 5
30 fragments obtained by PCR with the primers of set V are mixed, in tube 6 fragments obtained by PCR with the primers of set VI are mixed.
     Then, a SOE-PCR is performed by mixing aliquots of tube 1 and 2 and using the resulting mixture as template for a primary
35 SOE-PCR with corresponding external primers. The same is performed with mixtures of aliquots of tubes 3 and 4 as well as tubes 5 and 6. The respective external primer pairs are TiK134.#/125.# for fragments 1 and 2, TiK126.#/129.# for fragments 3 and 4, and TiK

130.#/135.# for fragments 5 and 6. The amplified assembled frag-
ments of about 340, 260, and 280 bp length, respectively, are pu-
rified by agarose gel electrophoresis. In a secondary SOE-PCR the
obtained fragments are mixed and assembled using primer pair
5 TiK134.#/135.# as external primers. The obtained full-length pro-
tease genes are gel-purified as described above.

    In another example, aliquots of tubes 1, 2, and 3 are mixed
and re-assembled by a primary SOE-PCR with primer pair
TiK134.#/127.#. Aliquots of tubes 4, 5, and 6 are also mixed in
10 another tube and re-assembled by another SOE-PCR using the primers
TiK128.#/135.#. The generated fragments of about 450 bp length are
purified as described above, mixed and reassembled in a secondary
SOE-PCR with external primers TiK134.#/135.#. The obtained full-
length protease genes are gel-purified as described above.

15   In principle, every combination of fragments may be assembled
in separate SOE-PCRs. In subsequent SOE-PCRs the obtained assem-
bled units are assembled to larger units until the final full
length gene is obtained. The overall number of SOE-PCRs used for
that purpose is only limited by experimental capacity. The only
20 prerequisite which is inherent to SOE-PCR is that the fragments to
be assembled must contain a sequence overlap as defined earlier.

    C) Cloning of the SOE-PCR-derived full-length protease-hybrids to
    yield library #1
25   The full-length protease-hybrid genes from step B2) as well
as the newly constructed shuttle vectors pTK4001-4005 from A3) are
separately digested with Acc65 I and Mlu I. In standard ligation
procedures the protease genes are separately ligated to each of
the five vectors pTK4001-4005 and transformed into E.coli DH10B.
30 Selection of correctly transformed cells is performed with am-
picillin. DNA of these clones is prepared and designated library
#1. The library size is about 10^5 independent transformants.

    D) Screening of library #1
35   Aliquots of library #1 are used to transform Bacilli cells
DN1885. The transformants are screened for the desired properties.

By this method and using a standard protease activity assay to screen for the desired property in step D) above a number of new shuffled subtilisins with a desired property were identified. The results are indicated in Table 1 below.

5

Table 1

Clone	pre-pro	frag.1 (5')	frag.2	frag.3	frag.4	frag.5	frag.6 (3')
8	BPN	Sav	Sav	Sav	Sav	Sav	Sav
6	Alc	Sav	Sav	Sav	Sav	Sav	Sav
12	Esp	Sav	Sav	Sav	Sav	Sav	Sav
10	PD498	Sav	Sav	Sav	Sav	Sav	Sav
4	Esp	PD138	Esp	Esp	Esp	Esp	JA16
22	Alc	PD138	Esp	Esp	Esp	Esp	JA16
11	PD498	PD138	Esp	Esp	Esp	Esp	JA16
1	Alc	PD138	Esp	PD138	Esp	Esp	JA16
3	BPN	PD138	Esp	Esp	PD138	Sav	Sav
17	Esp	PD138	PD138	Esp	Esp	Esp	JA16
19	PD498	Alc	BPN	Esp	Esp	Esp	JA16
16	Alc	Alc	BPN	Esp	PD138	Esp	JA16

Identity of clones:
10 Alcalase: A13050_1 (GenBank)   BPN': Poo782 (SwisProt)
   Esperase: D26542 (GenBank)   Savinase: A22550 (GenBank)
   PD498: WO 96/34963   JA16: WO 92/17576
   PD138 WO 93/18140

15   23 clones having protease activity were identified of which 12 were different. Clones 8, 9, 18, 20, 23 were the same; clones 6, 15, 21 were the same, clones 12, 14 were the same, clones 10, 13 were the same, and clones 4, 7 were the same. In respect of mature enzymes 7 different were identified.

20   From Table 1 it is seen that the process of the invention makes it possible to obtain active proteins representing combinations of proteins quite distantly related.

25 Example 2

The same methods as described in example 1 can be used for amplification of PCR fragments from fungal lipases.

The fungal lipases from the following fungi are aligned using the alignment program from Geneworks (using the following pa-

rameters:cost to open a gap = 5, cost to lengthen a gap = 25,
Minimum Diagonal lLength = 4, Maximum Diagonal Length = 10, Con-
sensus cutoff = 50%): Rhizomucor Miehei (LIP_RHIMI from the Swiss
Prot data base), Rhizopus Delemar (LIP_RHIDL from the Swiss Prot
5 data base), Penecillium camenbertii (MDLA_PENCA from the Swiss
Prot data base) Absidia reflexa (WO 96/13578) and Humicola lanug-
inosa (US 5536661).

Primers for amplification of Absidia (Absidia), Rhizopus
(LIP_RHIDL) and Rhizomucor(LIP_RHIMI) lipase genes for shuffling
10 *N: according to the IUPAC nomenclature means all 4 bases
(A,T,G,C).*

Set 1)
5' primer for YCRT/SVI/VPG: TAY TGY MGR ACN GTN ATH CCN GG or
15   TAY TGY MGR AGY/TCN GTN GTN CCN GG
3' primer for VFRGT/S: NSW NCC YCK RAA NAC

Set 2)
5' primer for VFRGT/S: GTN TTY MGR GGN WSN
20 3' primer for KVHK/AGF: RAA NCC YTT RTG NAC YTT or
                          RAA NCC NGC RTG NAC YTT

Set 3)
5' primer for KVHK/AGF: AAR GTN CAY AAR GGN TTY or
25   AAR GTN CAY GCN GGN TTY
3' primer for VTGHSLGG: CC NCC YAR NGA RTG NCC NGT NAC or
                        CC NCC YAR RCT RTG NCC NGT NAC

Set 4)
30 5' primer for VTGHSLGG: GTN ACN GGN CAY TCN YTR GGN GG or
                         GTN ACN GGN CAY AGY YTR GGN GG
3' primer for FGFLH: RTG YAR RAA NCC RAA

Set 5)
35 5' primer for FGFLH: TTY GGN TTY YTR CAY
3' primer for IVPFT: NGT RAA NGG NAC DAT

24

Primers for amplification of Humicola lanuginosa(Humicola) and
Penicillium camenbertii (MDLA_PENCA) lipase genes for shuffling

Set 1)
5 5' primer for CPEVE: TGY CCN GAR GTN GAR
   3' primer for VLS/AFRG: NCC YCK RAA NGM YAR NAC

   Set 2)
   5' primer for VLS/AFRG: GTN YTR KCN TTY MGR GGN
10 3' primer for GFT/WSSW: CCA NGA NGA NGT RAA NCC or
                           CCA RSW RSW CCA RAA NCC

   Set 3)
   5' primer for GFT/WSSW: GGN TTY ACN TCN TCN TGG or
15   GGN TTY TGG WSY WSY TGG
   3' primer for GHSLGG/AA: NGC NSC NCC YAR NGA RTG NCC or
                            NGC NSC NCC YAR RCT RTG NCC

   Set 4)
20 5' primer for GHSLGG/AA: GGN CAY TCN YTR GGN GSN GCN or
                            GGN CAY AGY YTR GGN GSN GCN
   3' primer for PRVGN: RTT NCC NAC YCK NGG

   Set 5)
25 5' primer for PRVGN: CCN MGR GTN GGN AAY
   3' primer for THTND: RTC RTT NGT RTG NGT

   Set 6)
   5' primer for THTND: ACN CAY ACN AAY GAY
30 3' primer for PEYWI: DAT CCA RTA YTC NGG

   Set 7)
   5' primer for PEYWI: CCN GAR TAY TGG ATH
35 3' primer for AHL/IWYF: RAA RTA CCA DAK RTG NGC

   Primers for shuffling of all five genes:

Set 1)

5' primer for AN/TA/SYCR: GCN AMY KCN TAY TGY MG for Absidia, Rhizopus and Rhizomucor sequences

5' primer for AN/TA/SYCGKNNDA: GCN AMY KCN TAY TGY GGN AAR AAY AAY

5 GAY GC for Humicola

5' primer for AN/TA/SYCEADYTA: GCN AMY KCN TAY TGY GAR GCN GAY TAY ACN GC for P. camenbertii

3' primer for E/QKTIY: RTA DAT NGT YTT YTS for Absidia, Rhizopus

10 and Rhizomucor sequences

3' primer for ALDNTE/QKTIY: RTA DAT NGT YTT YTS NGT RTT RTC YAR NGC for Humicola

3' primer for AVDHTE/QKTIY: RTA DAT NGT YTT YTS NGT RTG RTC NAC NGC for P. camenbertii

15

Set 2)

5' primer for E/QKTIY: SAR AAR ACN ATH TAY for Absidia, Rhizopus and Rhizomucor sequences

5' primer for E/QKTIYLA/SFRG: SAR AAR ACN ATH TAY YTR KCN TTY MGR

20 GGN for the two other sequences

3' primer for KVHK/AGF: RAA NCC YTT RTG NAC YTT or RAA NCC NGC RTG NAC YTT for Absidia, Rhizopus and Rhizomucor sequences

3' primer for ICSGCKVHK/AGF: RAA NCC YTT RTG NAC YTT RCA NCC NGA

25 RCA DAT or RAA NCC NGC RTG NAC YTT RCA NCC NGA RCA DAT for Humi-cola

3' primer for LCDGCKVHK/AGF: RAA NCC YTT RTG NAC YTT RCA NCC RTC RCA YAR or RAA NCC NGC RTG NAC YTT RCA NCC RTC RCA YAR for P. ca-menbertii

30

Set 3)

5' primer for KVHK/AGF: AAR GTN CAY AAR GGN TTY or AAR GTN CAY GCN GGN TTY for Absidia, Rhizopus and Rhizomucor sequences

5' primer for KVHK/AGFTSSW: AAR GTN CAY AAR GGN TTY ACN TCN TCN

35 TGG or AAR GTN CAY GCN GGN TTY ACN TCN TCN TGG for Humicola

5' primer for KVHK/AGFWSSW: AAR GTN CAY AAR GGN TTY TGG WSY WSY TGG or AAR GTN CAY GCN GGN TTY TGG WSY WSY TGG for P. camenbertii

3' primer for GHSLGG/AA: NGC NSC NCC YAR NGA RTG NCC or NGC NSC
NCC YAR RCT RTG NCC for all five sequences

Set 4)
5' primer for GHSLGG/AA: GGN CAY TCN YTN GGN GSN GCN or GGN CAY
AGY YTN GGN GSN GCN for all five sequences

3' primer for PRVGN/D: RTY NCC NAC YCK NGG for all the genes ex-
cept Absidia
3' primer for TQGQPRVGN/D: RTY NCC NAC YCK NGG YTG NCC YTG NGT for
Absidia

Set 5)
5' primer for PRVGN/D: CCN MGR GTN GGN RAY for all the genes ex-
cept Absidia
5' primer for PRVGN/DPAFA: CCN MGR GTN GGN RAY CCN GCN TTY GCN for
Absidia

3' primer for RDIVPH/R/K: YK NGG NAC DAT RTC YCK for Absidia,
Rhizopus and Rhizomucor sequences
3' primer for I/FTHTRDIVPH/R/K: YK NGG NAC DAT RTC YCK NGT RTG NGT
RAW for the two other sequences

Set 6)
5' primer for RDIVPH/R/K: MGR GAY ATH GTN CCN MR for Absidia,
Rhizopus and Rhizomucor sequences
5' primer for RDIVPH/R/KLP: MGR GAY ATH GTN CCN MRN YTR CCN for
the two other sequences

3' primer for EYWIK/T: YKT DAT CCA RTA YTC for Rhizomucor, Humi-
cola and P.camenbertii
3' primer for PGVEYWIK/T: YKT DAT CCA RTA YTC NAC NCC NGG for
Rhizopus
3' primer for AGEEYWIK/T: YKT DAT CCA RTA YTC YTC NCC NGC for Ab-
sidia

Set 7)

5' primer for EYWIK/T: GAR TAY TGG ATH AAR or GAR TAY TGG ATH ACN
for Rhizomucor, Humicola and P.camenbertii

5' primer for EYWIKSGT: GAR TAY TGG ATH AAR WSY GGN ACN for
5 Rhizopus

5' primer for EYWIKKDSS: GAR TAY TGG ATH AAR AAR GAY WSY WSY for
Absidia

3' primer for DHLSY: RTA NGA/RCT YAR RTG RTC for Absidia, Rhizopus
10 and Rhizomucor sequences

3' primer for IPDIPDHLSY: RTA NGA/RCT YAR RTG RTC NGG DAT RTC NGG
DAT for Humicola

3' primer for TDFEDHLSY: RTA NGA/RCT YAR RTG RTC YTC RAA RTC NGT
for P.camenbertii

15

        For the SOE-PCR the 5' primers from the first set of primers
and the 3' primer for the last set of primers can be used.

        The SOE-PCR fragments can then be combined with a lipase 5'
and 3' end, when the 5' and 3' ends have been generated by PCR.
20 The 5' end can be generated by PCR by using specific 5' primers
(containing a sequence for the BamHI recognition site in the 5'
end) for the 5' end of the genes of interest and using the comple-
mentary sequence from the 5' primer from the first set of primers
as the 3' primer. The 3' end can be generated by PCR by using spe-
25 cific 3' primers (containing a sequence for the XbaI recognition
site in the 5' end) for the 3' end of the genes of interest and
the complementary sequence from the 3' primer from the last set of
primers as the 5' primer.

        A second SOE is then used to generate the complete sequence,
30 by using the specific 5' and 3' primers from the genes of inter-
est.

        The genes can then be cloned into the yeast vector pJSO26 as
a BamHI-XbaI fragment (see WO 97/07205).

Example 3

        The overall same method as described in example 2 can be
used for amplification and recombination of PCR fragments of
Pseudomonas lipases. The term "overall same method" denotes that

it may be advantageous to use slightly different vectors as compared to example 2. Based on the sequence and primer information disclosed below it is a matter of routine for a person skilled in the art to modify the vectors etc. from example 2, in order to recombine below mentioned Pseudomonas lipases according to a shuffling method of the invention.

The Pseudomonas lipases mentioned below are aligned using the alignment program from Geneworks (using the following parameters:cost to open a gap = 5, cost to lengthen a gap = 25, Minimum Diagonal lLength = 4, Maximum Diagonal Length = 10, Consensus cutoff = 50%).

Pseudomonas lipases
Pseudomonas aeruginosa TE3285 (file ate3285d)
Pseudomonas pseudoalcaligenes M1 (Lipomax wt) (file pseudm1d)
Pseudomonas sp. SD705 (mature)(file spsd705d)
Pseudomonas wisconsinensis (file wisconsd) Proteus vulgaris KS0 (file provulgd) Pseudomonas fragi IFO 12049 (file fr12049d).

Suitable primers for shuffling of Pseudomonas lipases:
I = Inosin, Numbers refer to the numbers in the alignment(see figure 4), S means sense strand, the antisense oligonucleotide is of course also used:

5 109-131
   S1: 5'-TA(C/T)CCIAT(C/T)(G/T)I(C/T)T(G/A)(G/A)(C/T)ICA(C/T)GG-3'

   250-269
   S2: 5'-GA(G/A)(G/C)IICGIGGIG(A/C)I(G/C)A(G/A)(T/C)T-3'
10
   318-343
   S3: 5'-GT(C/A)AA(C/T)(C/T)T(G/A)ITCGG(C/T)CA(C/T)AG(C/T)CAIGG-3'

   607-628
15 S4: 5'-
   TIAA(C/T)(G/C/A)(G/C/A)(C/T/A)(A/C)(A/G)I(T/C)(A/T)(C/T)CCI(C/T)(A
   /G)(T/G/A)GG-3'

   801-817
20 S5: 5'-AA(C/T)GA(C/T)GG(C/T)(C/A/T)TGGT(C/T/G)GG-3'

   871-890
   S6: 5'-
   CA(C/T)(C/G)T(C/G)GA(C/T)(G/A)(A/C/T)(G/C)(G/A)T(G/C/A)AACCA-3'

CLAIMS

1. A method for shuffling of heterologous sequences of interest comprising the following steps,

5   i)   identification of at least one conserved region be-
                tween the heterologous sequences of interest;
         ii)   generating fragments of each of the heterologous se-
                quences of interest, wherein said fragments comprise
                the conserved region(s); and
10   iii)   shuffling/recombining said fragments using the con-
                served region(s) as (a) homologous linking point(s).

2. A method for producing a shuffled protein having a desired bio-
logical activity comprising in addition to the steps of the claim
15 1 the following further steps:
         iv)   expressing the numerous different recombinant proteins
                encoded by the numerous different shuffled sequences
                from step iii) (in claim 1); and
         v)   screen or select the numerous different recombinant
20   proteins from step ii) in a suitable screening or se-
                lection system for one or more recombinant protein(s)
                having a desired activity.

3.   The method for shuffling of heterologous DNA sequences of
25 interest, according to claim 1, having at least one conserved
region comprising the following steps
         i)   identification of one or more conserved region(s) (hereafter
                named "A,B,C" etc..) in two or more of the heterologous
                sequences;
30 ii)   construction of at least two sets of PCR primers (each set
                comprising a sense and an anti-sense primer) for one or more
                conserved region(s) identified in i) wherein
                in one set the sense primer (named: "a"=sense primer) is
                directed to a sequence region 5´ (sense strand) of said
35   conserved region (e.g. conserved region "A"), and the anti-
                sense primer (named "a´"=anti-sense primer) is directed
                either to a sequence region 3´ (sense strand) of said

conserved region or directed to a sequence region at least partially within said conserved region,

and in the second set the sense primer (named: "b"=sense primer) is directed either to a sequence region 5´ (sense strand) of said conserved region or directed to a sequence region at least partially within said conserved region and the anti-sense primer (named: "b´"=anti-sense primer) is directed to a sequence region 3´ (sense strand) of said conserved region (e.g. conserved region "A"), and

the two sequence regions defined by the regions between primer set "a" and "a´" and "b" and "b´" (both said regions is including the actual primer sequences) have a homologous sequence overlap of at least 10 base pairs (bp) within the conserved region;

iii) for one or more identified conserved region of interest in step i) two PCR amplification reactions are performed with the heterologous DNA sequences in step i) as template, and where

one of the PCR reactions is using the 5´ primer set identified in step ii) (e.g. named "a","a´") and the second PCR reaction is using the 3´ primer set identified in step ii) (e.g. named "b","b´");

iv) isolation of the PCR fragments generated as described in step iii) for one or more of the identified conserved region in step i);

v) pooling of two or more isolated PCR fragments from step iv) and performance of a Sequence overlap extension PCR reaction (SOE-PCR) using said isolated PCR fragments as templates; and

vi) isolation of the PCR fragment obtained in step v), wherein said isolated PCR fragment comprises numerous different shuffled sequences containing a shuffled mixture of the PCR fragments isolated in step iv), wherein said shuffled sequences are

characterized in that the partial DNA sequences, originating from the homologous sequence overlaps in step ii), have at least 80% identity to one or more partial sequences in one or more of the original heterologous DNA sequences in step i).

4. The method for producing one or more recombinant protein(s) having a desired biological activity, according to claim 2, comprising:

5 shuffling of heterologous DNA sequences, having at least one conserved region, encoding a protein by

i)   identification of one or more conserved region(s) (hereafter named "A,B,C" etc..) in two or more of the heterologous sequences;

10 ii)   construction of at least two sets of PCR primers (each set comprising a sense and an anti-sense primer) for one or more conserved region(s) identified in i) wherein

in one set the sense primer (named: "a"=sense primer) is directed to a sequence region 5´ (sense strand) of said

15   conserved region (e.g. conserved region "A"), and the anti-sense primer (named "a´"=anti-sense primer) is directed either to a sequence region 3´ (sense strand) of said conserved region or directed to a sequence region at least partially within said conserved region,

20   and in the second set the sense primer (named: "b"=sense primer) is directed either to a sequence region 5´ (sense strand) of said conserved region or directed to a sequence region at least partially within said conserved region and the anti-sense primer (named: "b´"=anti-sense primer) is

25   directed to a sequence region 3´ (sense strand) of said conserved region (e.g. conserved region "A"), and

the two sequence regions defined by the regions between primer set "a" and "a´" and "b" and "b´" (both said regions is including the actual primer sequences) have a homologous

30   sequence overlap of at least 10 base pairs (bp) within the conserved region;

iii)   for one or more identified conserved region of interest in step i) two PCR amplification reactions are performed with the heterologous DNA sequences in step i) as template, and

35   where

one of the PCR reactions is using the 5´ primer set identified in step ii) (e.g. named "a","a´") and the second

32

PCR reaction is using the 3` primer set identified in step
ii) (e.g. named "b","b´");

iv)   isolation of the PCR fragments generated as described in
       step iii) for one or more of the identified conserved region
5   in step i);

v)   pooling of two or more isolated PCR fragments from step iv)
       and performance of a Sequence overlap extension PCR reaction
       (SOE-PCR) using said isolated PCR fragments as templates;
       and

10 vi) isolation of the PCR fragment obtained in step v), wherein
       said isolated PCR fragment, comprises numerous different
       shuffled sequences containing a shuffled mixture of the PCR
       fragments isolated in step iv), wherein said shuffled
       sequences are

15 characterized in that the partial DNA sequences, originating from
   the homologous sequence overlaps in step ii), have at least 80%
   identity to one or more partial sequences in one or more of the
   original heterologous DNA sequences in step i);

20 vii) expressing the numerous different recombinant proteins
        encoded by the numerous different shuffled sequences in step
        vi); and

   viii) screen or select the numerous different recombinant proteins
        from step vii) in a suitable screening or selection system
25   for one or more recombinant protein(s) having a desired
        activity.

   5. The method according to any of claims 1-4, wherein the hete-
   rologous sequences of interest are encoding an enzyme.

30
   6. The method according to claim 5, wherein the enzyme is a
   protease, preferably a serine protease, and in particular a
   subtilase; or a lipase.

35 7. The method according to any of claims 3 and 4, wherein the PCR
   amplification process in step iii) is performed under conditions
   resulting in a low, medium or high random mutagenesis frequency.

8. The method according to any of claims 2 and 4, wherein the desired activity is an activity which leads to performance of the recombinant protein(s) in a dish-wash or laundry detergent.
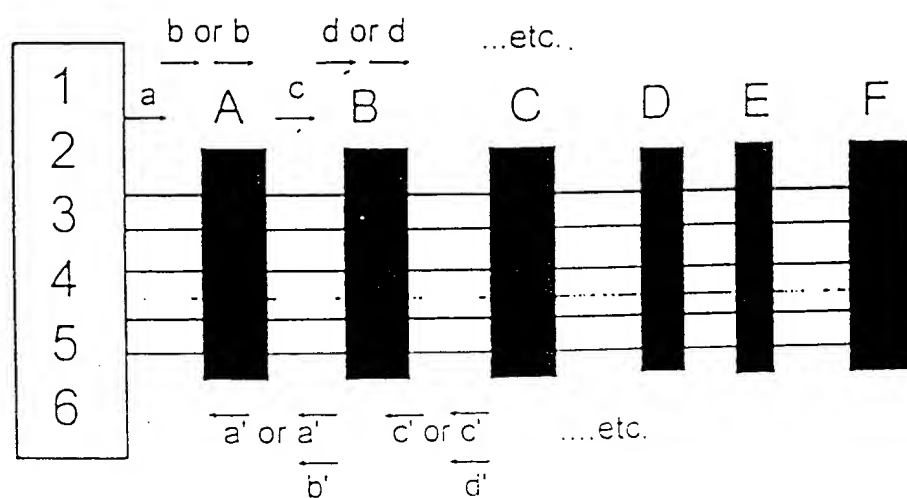
Fig. 1

```
  1  M M R - K K S F W L G M L T A F M L V F T M A F S D S A S A   A13050_1.PRO
  1  M - K - K P L G K I V A S T A L L I S - - V A F S S S I A S   A22550.PRO
  1  M - R - Q S L K V M V L S T - - - V A - - L L F M A N P A A   D26542.PRO
  1  M - R - G K K V W I S L L F A L A L I F T M A F G S T S S A   P00782.PRO
  1  M - K F K K I A A L S L A T S L A L F - - P A F G G S S L A   PD493.PRO

 30  A Q P - - - - - - - - - A K - N V E K - - - - - - Q - Y I V G F   A13050_1.PRO
 27  A A E - - - - - - - - - E A - - - - K - - - - - - E K Y L I - -   A22550.PRO
 24  A S E - - - - - - - - - E K - - - K - - - - - - E - Y L I - -   D26542.PRO
 29  Q A A - - - - - - - - - G K S N G E K - - - - - - K - Y I V G F   P00782.PRO
 28  K E A P K P F Q P I N K - T L D K G A F E S G E - V I V K F   PD493.PRO

 45  K S G - - - - - - - - - - - - - - - V K - - T - - A S V K K D   A13050_1.PRO
 38  - - - G F N E Q E A V S E F V E Q V E - - A N D E V A I L S   A22550.PRO
 34  - - - - - - - - - - - - - - - - - - V V - - E P E E V S A Q S   D26542.PRO
 45  K Q T - - - - - - - - - - - - - - - M S - - T M S A A K K K D   P00782.PRO
 56  K D G - - - - - - - - - - - - - - - V S K K A Q G S A L N K A   PD493.PRO

 57  I I K E S G G K V Q K Q F R I I N A A K A K L D K E A L K E   A13050_1.PRO
 63  E E E V E I E L L H E F E T I P V L S V E L S P E D V D A   A22550.PRO
 45  V E E S Y D V D V I H E F E I P V I H A E L T K K E L K K   D26542.PRO
 59  V I S E K G G K V Q K Q F K Y V D A A S A T L N E K A V K E   P00782.PRO
 72  E A N E Q K A S A K D P F Q V L E V A D V - - - D Q A V K A   PD493.PRO

 87  V K N D P D V A Y V E E D - - - - - - H V A H A L A Q T V   A13050_1.PRO
 93  L E L D P A I S Y I E E D - - - - - - A E V T M A Q S V   A22550.PRO
 75  L K K D P N V K A I E E N - - - - - - A E V T - I S Q T V   D26542.PRO
 89  L K K D P S V A Y V E E D - - - - - - H V A H A Y A Q S V   P00782.PRO
 99  L E N P N V E Y A F P N Y T F Q A T W S P N D P Y Y S A Y   PD493.PRO

110  P Y G I P L I K A D K V Q A Q G F K G A N V K V A V L D T G   A13050_1.PRO
116  P W G I S R V Q A P A A H N R G L T G S G V K V A V L D T G   A22550.PRO
 97  P W G I S F I N T Q Q A H N R G I F G N A R V A V L D T G   D26542.PRO
112  P Y G V S Q I K A P A L H S Q G Y T G S N V K V A V I D S G   P00782.PRO
129  Q Y G P Q N T S T P A A W D V T R G S S T Q T V A V L D S G   PD493.PRO

140  I Q A S H P D L - - N V V G G A S F V A G E A Y N - T D G N   A13050_1.PRO
146  I - S T H P D L - - N I R G G A S F V P G E P S T - Q D G N   A22550.PRO
127  I - A S H P D L - - R I A G G A S F I S S E P S Y - H D N N   D26542.PRO
142  I D S S H P D L - - K V A G G A S M V P S E T N P F Q D N N   P00782.PRO
159  V D Y N H P D L A R K V I K G Y D F I D R D N N P - M D L N   PD493.PRO

167  G H G T H V A G T V A A - L D N T T G V L G V A P S V S L Y   A13050_1.PRO
172  G H G T H V A G T I A A - L N N S I G V L G V A P S A E L Y   A22550.PRO
153  G H G T H V A G T I A A - L N N S I G V L G V A P S A D L Y   D26542.PRO
170  S H G T H V A G T V A A - L N N S I G V L G V A P S A S L Y   P00782.PRO
188  G H G T H V A G T V A A D T N N G I G V A G M A P D T K I L   PD493.PRO

196  A V K V L N S S G S G S Y S G I V S G I E W A T T N G M D V   A13050_1.PRO
201  A V K V L G A S G S G S V S S I A Q G L E W A G N N G M H V   A22550.PRO
182  A V K V L D R N G S G S L A S V A Q G I E W A I N N N M H I   D26542.PRO
199  A V K V L G A D G S G Q Y S W I I N G I E W A I A N N M D V   P00782.PRO
213  A V R V L D A N G S G S L D S I A S G I R Y A A D Q G A K V   PD493.PRO
```

Fig. 2 (a)

```
226  I N M S L G G A S G S T A M K Q A V D N A Y A R G V V V V A    A13050_1.PRO
231  A N L G L G S P S P S A T L E Q A V N S A T S R G V L V V A    A22550.PRO
212  I N M S L G S T S G S S T L E L A V N R A N N A G I L L V G    O26542.PRO
229  I N M S L G G P S G S A A L K A A V D K A V A S G V V V V A    P00732.PRO
248  L N L S L G C E C N S T T L K S A V O Y A W N K G A V V V A    P0493.PRO


256  A A G N S G S S G N T N T I G Y P A K Y D S V I A V G A V D    A13050_1.PRO
261  A S G N S G A G S I S - - - - Y P A R Y A N A M A V G A T D    A22550.PRO
242  A A G N T G R Q G V N - - - - Y P A R Y S G I V M A V A A V D  O26542.PRO
259  A A G N E G T S G S S S T V G Y P G K Y P S V I A V G A V D    P00732.PRO
278  A A G N O N V S R T F - - - - Q P A S Y P N A I A V G A I D    P0493.PRO


296  S N S N R A S F S S V G A E L E V M A P G A G V Y S T Y P T    A13050_1.PRO
297  Q N N N R A S F S Q Y G A G L D I V A P G V N V Q S T Y P G    A22550.PRO
258  Q N G Q R A S F S T Y G P E I E S A P G V N V N S T Y T G      O26542.PRO
289  S S N Q R A S F S S V G P E L D V M A P G V S I Q S T L P G    P00732.PRO
304  S N D R K A S F S N Y G T W V O V T A P G V N I A S T V P N    P0493.PRO


316  N T Y A T L N G T S M A S P H V A G A A A L I L S K H P N L    A13050_1.PRO
317  S T Y A S L N G T S M A T P H V A G A A A L V K Q K N P S W    A22550.PRO
299  N R Y V S L S G T S M A T P H V A G V A A L V K S R Y P S Y    O26542.PRO
319  N K Y G A Y N G T S M A S P H V A G A A A L I L S K H P N W    P00732.PRO
334  N G Y S Y M S G T S M A S P H V A G L A A L L A S Q - - G K    P0493.PRO


346  S A S Q V R N P L S S T A T Y L - - - G S S F Y Y G K G L I    A13050_1.PRO
347  S N V Q I R N H L K N T A T S L - - - G S T N L Y G S G L V    A22550.PRO
329  T N Q I R Q R I N Q T A T Y L - - - G S P S L Y G N G L V      O26542.PRO
349  T N T Q V R S S L E N T T K L - - - G D S F Y Y G K G L I      P00732.PRO
362  N N V Q I R Q A I E Q T A D W S G T G T N F K Y G K - - I      P0493.PRO


373  N V E A A A Q                                    A13050_1.PRO
374  N A E A A T R                                    A22550.PRO
355  H A G R A T Q                                    O26542.PRO
375  N V Q A A A Q                                    P00732.PRO
390  N S N K A V R Y                                  P0493.PRO
```

Fig. 2 (b)

Percent Similarity

	1	2	3	4	5		
1	■	52.2	48.6	66.5	41.8	1	A13050_1.PRO
2	74.2	■	59.9	51.6	41.8	2	A22550.PRO
3	83.6	56.7	■	48.1	39.4	3	D26542.PRO
4	44.2	75.6	85.2	■	45.4	4	P00782.PRO
5	100.0	100.0	100.0	93.1	■	5	PD498.PRO
	1	2	3	4	5		

Fig. 2 (c)

```
LIP_RHIMI    MV-LKQRANY LGF-LIVFFT AFLVEAVPI- -KRQSNSTV- -----DSLPP   40
LIP_RHIDL    MVSFISISQG VSLCLLVSSH HLGSSAVPVS GKSGSSNTAV SASDNAALPP   50
ABSIDIA      M-----HSHF VVLLLAVFIC MCSVSGVPL- -QIDPRDDK- -----SYVPE   37
MDLA_PENCA   M--------R LSFFTALSA- ---VASLG-- ---------- ----YA-LPG   21
Humicola     M--------R SSL--VLFF- ---VSAWT-- ---------- -----A-LAS   18
Consensus    M-----.... .SL.L.VF.. ...VSAVP.- -.........- -----A.LP.   50


LIP_RHIMI    LIPSRTSAPS SSPSTTDPEA -P-AM----- ---SRNGPLP S--DVETKY-   77
LIP_RHIDL    LISSRCAPPS NKGSKSDLQA EPYNHQKNTE WYESHGGNLT SIGKRDDNLV  100
ABSIDIA      QYPLKVNGPL PEGVSVIQGY ---------- ---CENCTHY P----EKN--   68
MDLA_PENCA   KLQSR----- ---------- ---------- ---------- -----D----   27
Humicola     PIR-R----- ---------- ---------- ---------- -----E----   23
Consensus    .I.SR...P. .......... ---------- ---........ .----E..--  100


LIP_RHIMI    -GMALNATSY PDSVVQAMSI -----DCG-IR AATSQEINEL TYYTTLSANS  121
LIP_RHIDL    GGHTLDLPSD APPISLSSST NSASDGGKVV AATTAQIQEF TKYAGIAATA  150
ABSIDIA      -----SVSAF.SSSSTQDYRI ---------- -ASEAEIKAH TFYTALSANA  102
MDLA_PENCA   ---------- ---------- ---------- -VSTSELDQF EFWVQYAAAS   46
Humicola     ---------- ---------- ---------- -VSQDLFNQF NLFAQYSAAA   42
Consensus    -----..... .......... ---------- -AS..EI..F T.Y...SA.A  150


LIP_RHIMI    YC---RTVIP GATWDCI--H C-DA-TEDLK IIKTWS-TLI YDTNAHVARG  163
LIP_RHIDL    YC---RSVVP GNKWDCV--Q C-QKWVPDGK IITTFT-SLL SDTNGYVLRS  193
ABSIDIA      YC---RTVIP GGRWSCP--H C-GV-ASHLQ ITKTFS-TLI TDTNVLVAVG  144
MDLA_PENCA   YYEADYTAQV GDKLSCSKGN CPEVEATGAT VSYDFSDSTI TDTAGYIAVD   96
Humicola     YCGKNNDAPA GTNITCTGNA CPEVEKADAT FLYSFEDSGV GDVTGFLALD   92
Consensus    YC---RTV.P G..W.C.--. C-.V...D.. I..TFS-SLI .DTNG.VA..  200


LIP_RHIMI    DSEKTIYIVF RGSSSIRNWI ADLTFVPVSY PPV-SGTKVH KGFLDSYGEV  212
LIP_RHIDL    DKQKTIYLVF RGTNSFRSAI TDIVFNFSDY KPV-KGAKVH AGFLSSYEQV  242
ABSIDIA      EREKTIYVVF RGTSSIRNAI ADIVFVPVNY PPV-NGAKVH KGFLDSYNEV  193
MDLA_PENCA   HTNSAVVLAF RGSYSVRNWV ADATF-VHTN PGLCDGCLAE LGFWSSWRLV  145
Humicola     NTNKLIVLSF RGSRSIENWI GNLNFDLKEI NDICSGCRGH DGFTSSWRSV  142
Consensus    ...KTIYLVF RGS.SIRNWI AD..F....Y PPV-.G.KVH .GFLSSY..V  250


LIP_RHIMI    QNELVATVLD QFKQYPSYKV AVTGHSLGGA TALLCALDLY QREEGLSSSN  262
LIP_RHIDL    VNDYFPVVQE QLTAHPTYRV IVTGHSLGGA QALLAGHDLY QREPPLSPKN  292
ABSIDIA      QDRLVAEVKA QLDRHPGYKI VVTGHSLGGA TAVLSALDLY HHGH----AN  239
MDLA_PENCA   PDDIIKELRE VVAQHPNYEL VVTGHSLGAA VATLAATDLR GKGYP----S  191
Humicola     ADTLRQKVED AVREHPDYRV VFTGHSLGGA LATVAGADLR GNGY-----D  187
Consensus    .D.L...V.. Q...HP.YKV VVTGHSLGGA .A.LAA.DLY ..G..---.N  300


LIP_RHIMI    LFLYTQGQPR VGDPAFANYV VST-GIPYRR TVNERDIVPH LPPAAFGFLH  311
LIP_RHIDL    LSIFTVGGPR VGNPTFAYYV EST-GIPFQR TVHKRDIVPH VPPQSFGFLH  341
ABSIDIA      IEIYTQGQPR IGTPAFANYV IGT-KIPYQR LVHERDIVPH LPPGAFGFLH  283
MDLA_PENCA   AKLYAYASPR VGNAALAKYI TAQ--GNNFR FTHTNDPVPK LPLLSMGYVH  239
Humicola     IDVFSYGAPR VGNRAFAEFL TVQTGGTLYR ITHTNDIVPR LPPREFGYSH  237
Consensus    ...YT.G.PR VGNPAFA.YV ..T-GIP..R .VH.RDIVPH LPP..FGFLH  350
```

<div style="text-align:center">

Fig. 3 (a)

SUBSTITUTE SHEET (RULE 26)

</div>

```
LIP_RHIMI    AGEEYWITDN SPETVQVC-T SDLET----S DCSNSIVP-F TSVLDHLSYF   355
LIP_RHIDL    PGVESWIKSG TSN-VQIC-T SEIET----K DCSNSIVP-F TSILDHLSYF   384
ABSIDIA      AGEEFWIMKD SSLRV--C-P NGIET----D NCSNSIVP-F TSVIDHLSYL   330
MDLA_PENCA   VSPEYWITSP NNATVSTSDI KVIDGDVSFD GNTGTGLPLL TDFEAHIWYF   289
  Humicola   SSPEYWIKSG TLVPVTRNDI VKIEG---ID ATGGNNQPNI PDIPAHLWYF   284
Consensus    .G.EYWI.S. ....V..C-. ..IET----D .CSNSIVP-F TS..DHLSYF   400


LIP_RHIMI    GIN---TGLC T----- 363
LIP_RHIDL    DIN---EGSC L----- 392
ABSIDIA      DMN---TGLC L----- 338
MDLA_PENCA   VQVDAGKGPG LPFKRV 305
  Humicola   GLI----GTC L----- 291
Consensus    ..N---.G.C L----- 416
```

Fig. 3 (b)

```
ABSIDIA    M-----HSHF VVLLLAVFIC MCSVSGVPL- ---------- ------QIDP    28
LIP_RHIMI  MV-LKQRANY LGF-LIVFFT AFLVEAVPI- -KRQSNSTV- -----DSLPP    40
LIP_RHIDL  MVSFISISQG VSLCLLVSSM MLCJSAVPVS GKSGSSNTAV SASDNAALPP    50
Consensus  MV-....S.. V.L.L.VF.. M..VSAVP.- -K..S..T.- -----..LPP    50
                 .              . .                . .

ABSIDIA    -RDDKSYVPE QYPLKVN--- ---------- ------GPLP EGVSVIQGYC    53
LIP_RHIMI  LIPSRTSAPS SSPSTTDPEA -P-AM----- ---SRNGPLP S--DVETKY-    77
LIP_RHIDL  LISSRCAPPS NKGSKSDLQA EPYNMQKNTE WYESHGGNLT SIGKRDDNLV   100
Consensus  LI.SR...PS ..PSK.D..A -P-.M----- ---S..GPLP S...V...Y.   100
                 .              . .   .

ABSIDIA    ENCTMYPEKN SVSAFSSSST ----QD--YR IASEAEIKAH TFYTALSANA   102
LIP_RHIMI  -GMALNATSY PDSVVQAMSI ----DGG-IR AATSQEINEL TYYTTLSANS   121
LIP_RHIDL  GGMTLDLPSD APPISLSSST NSASDGGKVV AATTAQIQEF TKYAGIAATA   150
Consensus  .GMTL...S. ..S...SSST ----DGG-.R AAT.AEI.E. T.YT.LSANA   150
                 .              . .         . .

ABSIDIA    YCRTVIPGGR WSCPHCGV-A SMLQITKTFS TLITDTWVLV AVGEKEKTIY   151
LIP_RHIMI  YCRTVIPGAT WDCIHCDA-T EDLKIIKTWS TLIYDTNAMV ARGDSEKTIY   170
LIP_RHIDL  YCRSVVPGNX WDCVQCQKWV PGGKIITTFT SLLSDTNGYV LRSDKQKTIY   200
Consensus  YCRTVIPG.. WDC.HC..-. .DLKIIKTFS TLI.DTN..V ARGDKEKTIY   200
                 .              . .         . .

ABSIDIA    VVFRGTSSIR NAIADIVFVP WMYPPVNGAK VHKGFLDSYN EVQDKLVAEV   201
LIP_RHIMI  IVFRGSSSIR NWIADLTFVP VSYPPVSGTK VHKGFLDSYG EVQNELVATV   220
LIP_RHIDL  LVFRGTNSFR SAITDIVFNF SDYKPVKGAK VHAGFLSSYE QVVNDYFPVV   250
Consensus  .VFRGTSSIR NAIADIVFVP V.YPPV.GAK VHKGFLDSY. EVQN.LVA.V   250
                 .              . .         . .

ABSIDIA    KAQLDRHPGY KIVVTGHSLG GATAVLSALD LYHHGH---- ANIEIYTQGQ   247
LIP_RHIMI  LDQFKQYPSY KVAVTGHSLG GATALLCALD LYQREEGLSS SNLFLYTQGQ   270
LIP_RHIDL  QEQLTAHPTY KVIVTGHSLG GAQALLAGHD LYQREPRLSP KNLSIFTVGG   300
Consensus  ..QL..HP.Y KV.VTGHSLG GATALL.ALD LYQRE..LS. .NL.IYTQGQ   300
                 .              . .         . .

ABSIDIA    PRIGTPAFAN YVIGTKIPYQ RLVHERDIVP HLPPGAFGFL HAGEEFWIMX   297
LIP_RHIMI  PRVCDPAFAN YVVSTGIPYR RTVHERDIVP HLPPAAFGFL HAGEEYWITD   320
LIP_RHIDL  PRVGNPTFAY YVESTGIPFQ RTVHKRDIVP HVPPQSFGFL HPGVESWIKS   350
Consensus  PRVG.PAFAN YV.STGIPYQ RTVHERDIVP HLPP.AFGFL HAGEE.WI..   350
                 .              . .         . .

ABSIDIA    DSSLRV--CP NGIETDNCSN SIVPFTSVID HLSYLDMNTG LCL   338
LIP_RHIMI  NSPETVQVCT SDLETSDCSN SIVPFTSVLD HLSYFGINTG LCT   363
LIP_RHIDL  GTSN-VQICT SEIETKDCSN SIVPFTSILD HLSYFDINEG SCL   392
Consensus  .SS..VQ.CT S.IET.DCSN SIVPFTSVLD HLSYFDINTG LCL   393
                 .              . .         . .
```

Fig. 3 (c)

```
Humicola      MRSSL--VLF FVSAWT-ALA SPIR-REVSQ DLFNQFNLFA QYSAAAYCGK       46
MDLA_PENCA    MRLSFFTALS AVASLGYALP GKLQSRDVST SELDQFEFWV QYAAASYYEA       50
Consensus     MR.S....L. .V.....AL. .....R.VS. ....QF.... QY.AA.Y...       50
              ..  .      .       ..      . ..     ..    .. . .          .


Humicola      NNDAPAGTNI TCTGNACPEV EKADATFLYS FEDSGVGDVT GFLALDNTNK       96
MDLA_PENCA    DYTAQVGDKL SCSKGNCPEV EATGATVSYD FSDSTITDTA GYIAVDHTNS      100
Consensus     ...A..G... .C....CPEV E...AT..Y. F.DS...D.. G..A.D.TN.      100
              ..  .  .   ..     ..     . ..     ..    ..  . .  .


Humicola      LIVLSFRGSR SIENWIGNLN FDLREINDIC SGCRGHDGFT SSWRSVADTL      146
MDLA_PENCA    AVVLAFRGSY SVRNWVADAT F-VHTNPGLC DGCLAELGFW SSWKLVRDDI      149
Consensus     ..VL.FRGS. S..NW..... F........C .GC....GF. SSW..V.D..      150
              ..  .    . .  ..       F....     .. .  . . .. .   . .


Humicola      RQKVEDAVRE HPDYRVVFTG HSLGGALATV AGADLRGNGY -DIDVFSYGA      195
MDLA_PENCA    IKELKEVVAQ NPNYELVVVG HSLGAAVATL AATDLRGKGY PSAKLYAYAS      199
Consensus     .......V.. .P.Y..V..G HSLG.A.AT. A..DLRG.GY .......Y..      200
              .. .      .P. .  ..    ..    ..   ..    ..    .  .   ..


Humicola      PRVGNPAFAE FLTVQTGGTL YRITHTNDIV PRLPPPEFGY SHSSPEYWIK      245
MDLA_PENCA    PRVGNAALAK YITAQ--GNN FRFTHTNDPV PKLPLLSHGY VHVSPEYWIT      247
Consensus     PRVGN.A.A. ..T.Q..G.. .R.THTND.V P.LP....GY .H.SPEYWI.      250
              .. .  . .   ..    ..   .  .       .. . .       ..    .


Humicola      SGTLVPVTRN DIVRIEG--- IDATGGNNQP NIPDIPAHLW YFGLI----G      286
MDLA_PENCA    SPINATVSTS DIKVIDGDVS FDGNTGTGLP LLTDFEAHIW YFVQVDAGKG      297
Consensus     S.....V... DI..I.G... .D...G...P ...D..AH.W YF.......G      300
              S.    .     DI..I.G     .   .G   P    .D  AH.W YF       .G


Humicola      TCL-----   291
MDLA_PENCA    PGLPFKRV   305
Consensus     ..L.....   308
              .. .
```

Fig. 3 (d)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/DK 98/00105

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: C12N 15/10, C12Q 1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: C12N, C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 9522625 A1 (AFFYMAX TECHNOLOGIES N.V.), 24 August 1995 (24.08.95), page 9, line 6 - line 8; page 78, line 16 - line 25, claim 12	1-2,5-6
A		3-4,7-8

☐ Further documents are listed in the continuation of Box C.   ☒ See patent family annex.

* Special categories of cited documents

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 July 1998	03 -07- 1998

Name and mailing address of the ISA/	Authorized officer
Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Patrick Andersson Telephone No. +46 8 782 25 00

Form PCT/ISA/210 (second sheet) (July 1992)

Information on patent family members

09/06/98

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9522625 A1	24/08/95	AU	2971495 A	04/09/95
		CA	2182393 A	24/08/95
		CN	1145641 A	19/03/97
		EP	0752008 A	08/01/97
		JP	10500561 T	20/01/98
		US	5605793 A	25/02/97